ABSTRACT OF THE DISCLOSURE

An object of the present invention is to perform the clustering and assembling of nucleic acid base sequences at a high speed. Partial sequences 102 are extracted from each input sequence 101 and entered into a fixed-length partial sequence table 103. In the case where a sequence overlapping with a consensus sequence 104 is searched while making reference to the fixed-length partial sequence table 103 and consequently a partial sequence 102, which exactly matches with a sequence defined by a fixed length window 105 scanning along the consensus sequence, is found to be present, whether the whole input sequence can be assembled or not is determined by comparing the sequences. If it is possible to assemble the sequences, they are assembled into a consensus sequence and also joined into the same cluster. The clustering and assembling are performed by repeatedly processing this procedure based on greedy method until no unprocessed input nucleic acid base sequence is left.